ERDC MSRC PET Technical Report No. 01-19

Audio Video Conferencing

by

Geoffrey Fox
Gurhan Gunduz
Ahmet Uyar

14 May 2001

Work funded by the Department of Defense
High Performance Computing Modernization Program
U.S. Army Engineer Research and Development Center
Major Shared Resource Center through

Programming Environment and Training

Supported by Contract Number: DAHC94-96-C0002
Computer Sciences Corporation

Audio Video Conferencing

Geoffrey Fox, Gurhan Gunduz and Ahmet Uyar
Florida State University
Department of Computer Science and CSIT
(School of Computational Science and Information Technology)
400 Dirac Science Library
Tallahassee, Florida 32306-4120
fox@csit.fsu.edu

1 Introduction

This report describes in detail a focused effort studying the Access Grid and HearMe audio video systems. The original Tango Interactive system had its own audio video conferencing system, Buena Vista, built in. This is considered important for two reasons. First, it allowed a single invocation and user registration process for all parts of the collaboration; secondly, the alternatives developed by Buena Vista some four years ago were not very satisfactory. Garnet [1] was chosen because it is no longer realistic or even useful to develop audio video support. Instead, the best-of-practice solutions from other commercial or research developers will be used. In some cases, an available API to link invocation and registration may be used; in others, audio video and shared document support may be viewed as separate stand-alone systems. Audio video support is classified three ways:

1. **Low-end:** Illustrated by HearMe Audio and CUSeeMe Desktop Video;
2. **Medium:** Illustrated by PictureTel and Polycomm;
3. **High-end:** Illustrated by Access Grid and the system Admire from BUAA University in Beijing, China.

The Access Grid and HearMe systems are discussed in this report, and the remainder of the introduction describes general plans in this regard. The final two sections describe these systems in technical detail.

The Access Grid (AG) was originally developed by Argonne [3] but was extended as part of the NCSA Alliance. There are now more than 50 of these high-end audio video conferencing systems installed worldwide. Two systems are being installed at FSU – one fixed and one transportable model aimed for “teachers” in a distance education scenario or for use in a small conference room. The Access Grid is described at http://www.mcs.anl.gov/fl/accessgrid/ and training sessions are available. NCSA recently offered AG training to ERDC personnel at the Access Center in Washington, D.C., and training is expected for all the major PET and MSRC sites that wish to install AG nodes. A “train the trainers” model will also be established, placing a cadre of experts at HPCMO/PET sites that can help other sites in this community. This will build a critical mass of AG systems to enable effective electronic collaboration in HPCMO/PET. Currently, ERDC, JSU and ARL intend to install AG.

Although the AG is an impressive system, some issues need to be addressed. It is recommended that different shared document systems be used when the simple shared PowerPoint of the AG is insufficient. This motivates the system to use AG for community conferencing but Garnet or commercial collaboration systems such as Centra or Webex for document sharing. Further, the AG community should look at H.323 or SIP (the audio video interoperability standards) compliance for this technology. This would allow the user to support hybrid sessions involving simultaneous systems such as AG, PictureTel, Polycomm, CUSeeMe, and HearMe. Details of the H.323 and SIP standards are reviewed [4]. Finally, the AG needs to support partitioning of clients so that multiple communities can be administered separately. An important (for HPCMO) recent enhancement to the AG supports encrypted media streams using the new AES standard [5].

HearMe http://www.hearme.com [2] is a low-end audio conferencing system supporting general mix of phones and Internet clients with participant control. The phone option is helpful as it allows audio communication with better quality of service than can be guaranteed on the Internet. Note that the phone and Internet options are integrated as both are converted to the same codecs and recorded for later replay through the Web. SMIL-based replay (the W3C standard for multi-stream multi-media files) of session by converting G.711 or G.723 codecs digitized on the HearMe server to RealAudio has been added. A HearMe server with a license for 20 simultaneous users has been installed at FSU.

The Access Grid produces a “designed space” aimed at supporting groups interacting with groups; PictureTel or desktop systems are more optimized for individual interactions. The AG features hands-free, high-quality audio, multiple (four) video and audio streams, and lifesize displays. Four PCs control it, and AG equipment includes an echo canceling box, multiple camera, and projector or frame buffer displays (at least three).

HearMe provides two types of conferences: standard and moderated. In the standard conference, every client has the same privileges and can speak at any time. In a moderated conference, there are three types of users: moderator, panelist, and participant. The moderator is the creator of the conference and has full control over the session. Panelists are those given the right to speak, while participants can listen but need the moderator’s permission to speak.

2 HearMe voice over IP system

URL: http://www.hearme.com/

2.1 Introduction

HearMe [2] is a voice over IP application for voice conferencing. It provides full-duplex voice communication among participants. **It has no video capability.**

Today, there are three methods of teleconferencing: the Internet, through which people attend conferences by using PCs; phone lines with a conference typically arranged by telephone companies; and a combination of the Internet and phone lines. In this case, people can attend conferences by using either PCs or phones. The HearMe system is based on the third method.

Although using just the Internet for teleconferencing is inexpensive, the quality of voice is often unsatisfactory. On the other hand, the quality of is voice usually reasonable when phone lines are used. However, using phones is expensive and inconvenient for many people. The third solution combines the quality of phone lines and low cost of the Internet. The idea is that the speaker will talk on the phone, which provides better voice quality, and listeners can use either phones or PCs. In addition to its cost benefits, this solution is also more convenient than the other two. A phone-to-PC gateway is used to connect phone lines to the Internet.

2.2 Services

HearMe provides two types of conferences, standard and moderated. In a standard conference, everyone has the same privileges and can speak at any time. In a moderated conference, there are three types of users: moderator, panelist, and participant. The moderator has full control over the conference. He or she gives permission to speak and has the right to eject a participant from the conference. A panelist has the right to speak by default, but participants need permission to speak.

While HearMe provides a recording mechanism for live sessions, it unfortunately does not provide any tool to replay recorded conferences. Recorded conferences are in HearMe proprietary format and users need to write a decoder to replay it. FSU is currently implementing replay using the Internet standards such as RealPlayer or Microsoft technology.

2.3 Architecture

The three servers are TalkServer, MCU, and BridgeServer. TalkServer is a management tool that creates a conference, destroys a conference, gathers information about a conference, etc. TalkServer is basically used by administrators. The MCU (Multi-point control unit) does the real work, obtaining voice packages from different people and transmitting them to the appropriate recipients. In addition, MCU can record conferences. Users directly connect to the MCU, while BridgeServer and an IP gateway are used to include phone connections into conferences. Gateway converts analog voice signals to digital form and vice versa. BridgeServer is used as a bridge between the gateway and the MCU.

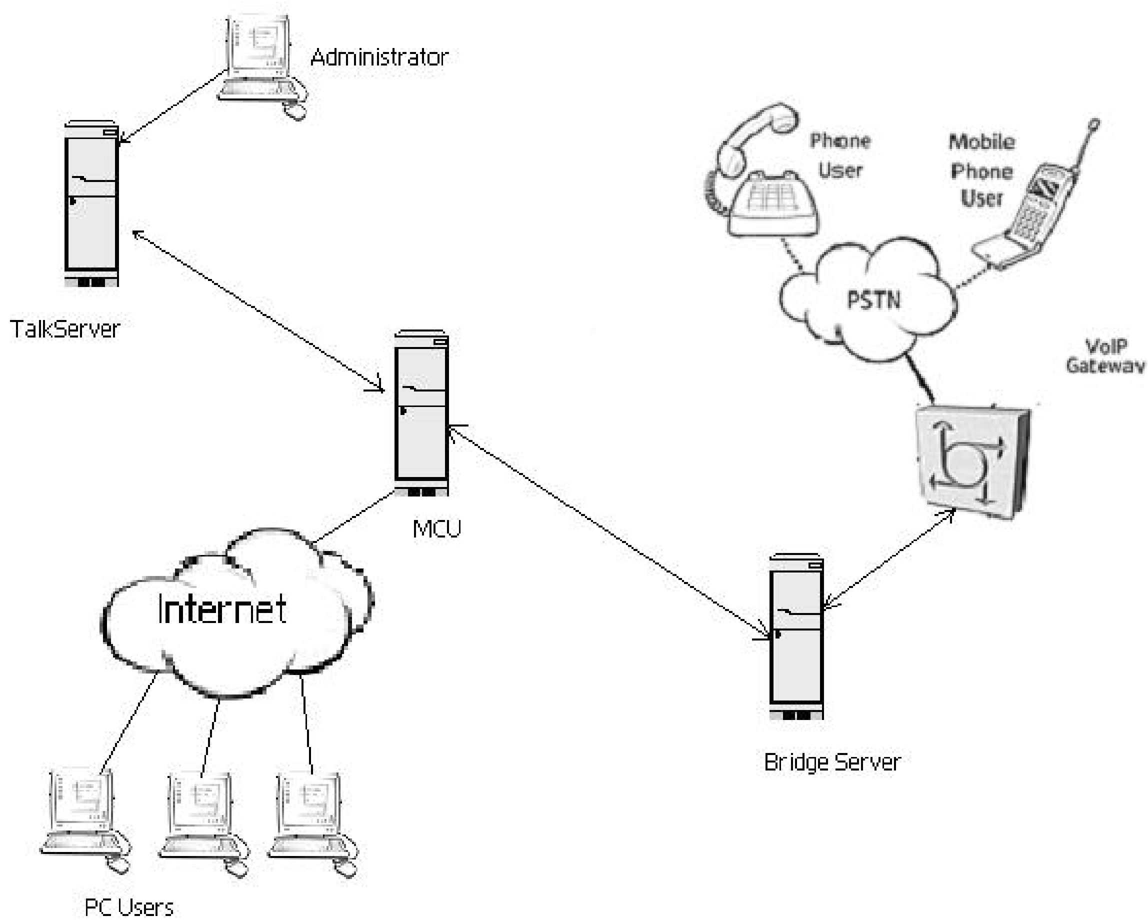


Figure 1: The architecture of HearMe voice over IP system.

2.4 Protocols

HearMe uses industry standards in its voice over IP system. The system architecture is based on the H.323 standard described in Ref. [4] that is recommended by International Telecommunication Union (ITU). It sets standards for multimedia communications over networks that do not provide quality of service, as well as sets standards for voice, video, and data. HearMe currently uses G.723.1 for voice compression. G.723.1 is also recommended by ITU and is widely used for Internet telephony and Web conferencing. ITU G.711 is also used for voice compression, which provides better voice quality and requires higher bandwidth, but it is currently not fully functional. In addition, HearMe uses session initiation protocol (SIP) to initiate sessions.

2.5 Bandwidth requirements

Each client needs 28.8 Kbps or greater Internet connection.

2.6 Client side system requirements

The minimum system requirements for each client are:

- Pentium 166MHz
- 32Mb of RAM
- Sound Blaster compatible 16-bit sound card
- Headset or speakers and microphone
- Windows 95, 98, or NT
- Internet Explorer 4.0 or later, or Netscape 4.5 or later

2.7 Server side System requirements

TalkServer:

- Pentium III @ 500MHz
- 256 MB RAM
- 10 GB disk
- 100 Mbit/sec network interface card
- RedHat Linux 6.1
- Oracle 8i

MCU:

- Pentium III @ 500MHz
- 256 MB RAM
- 10 GB disk
- 100 Mbit/sec network interface card
- RedHat Linux 6.1

BridgeServer:

- Pentium III @ 500MHz
- 256 MB RAM
- 10 GB disk
- 100 Mbit/sec network interface card
- RedHat Linux 6.1
- H.323 VoIP Gateway (ref.:Cisco AS5300)

2.8 Cost

The cost of a HearMe Voice Developer's Kit is $10,000. It includes:

- Server software for TalkServer, MCU, and BridgeServer.
- License files to allow service for up to 16 concurrent customers. More can be added at additional cost.
- HearMe Voice SDKs

2.9 Conclusion

HearMe provides a solution for the voice conferencing over the Internet and also allows telephone users to attend these conferences. It is relatively inexpensive and high quality compared with other solutions available on today's market. Although it lacks some features such as replaying recorded conferences, HearMe is on the right track and plans to add those features in future releases.

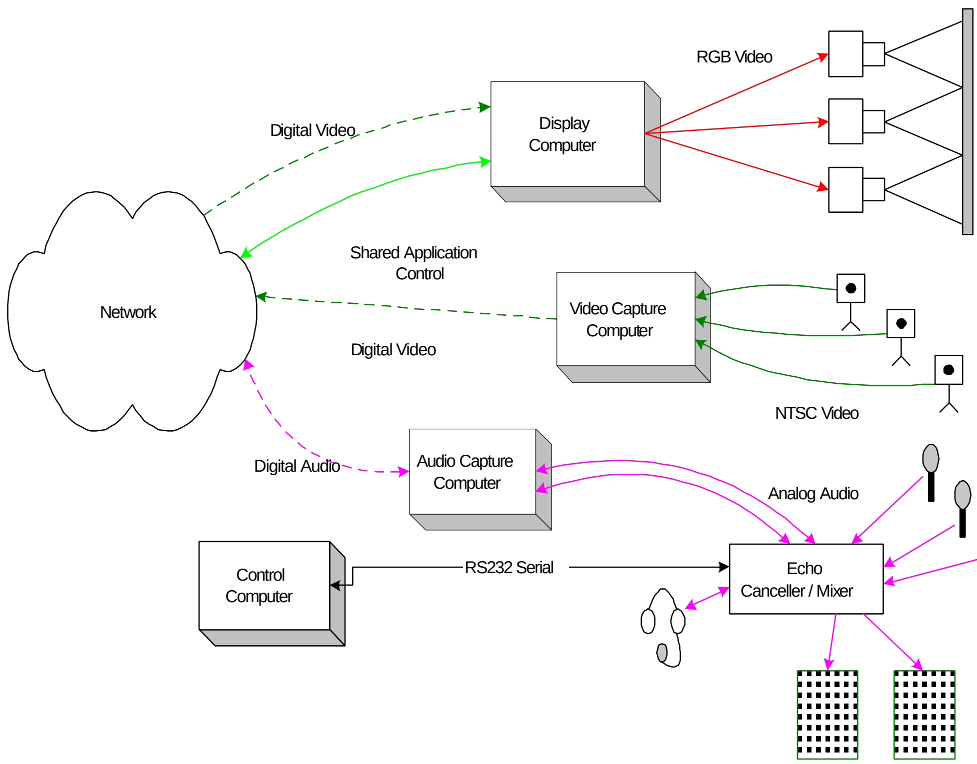
3 Access Grid

URL: http://www-fp.mcs.anl.gov/fl/accessgrid/default.htm

3.1 Introduction

The Access Grid [3], designed by Argonne National Laboratory, is a system that enables group-to-group collaboration across the Internet by providing multiple video and audio streams. The Access Grid consists of about 50 AG nodes across the United States. An AG node is a special room designed to participate in AG meetings. It consists of video cameras, projectors, audio equipment, computing equipment, and high-speed Internet connection.

The Access Grid project's focus is to enable *groups* of people to interact with grid resources and to use the grid technology to support group-to-group collaboration at a distance. This is the main difference between desktop-based collaboration tools and the AG. The AG is designed to give a sense of presence to remote participants. AG nodes have large displays, multiple video, and audio streams. Audio system is designed so that every participant can talk hands-free.



3.2 Video

Because it is important to see every participant in remote sites, each AG node has four video cameras. One camera is used to get the presenter's video stream, the second is for display screen shots, and the remaining cameras are for audience shots. Camera placement should facilitate the feeling of eye contact.

3.3 Audio

The most important thing in audio configuration is to enable all participants to talk hands-free. Therefore, there should be an adequate number of microphones properly placed throughout the room. There must be also an echo canceller device in each AG node. Two speakers are used to project good quality of audio.

3.4 Projectors

Since it is important to get real, lifesize images of participants at remote sites, large display screens are used in each AG node. This is accomplished by using three high-resolution projectors. Each node gets four video streams from participating nodes, so there are many video streams coming to one node.

3.5 Computers

Each AG node has four computers: display computer, video capture computer, audio capture computer, and control computer. The display computer is used to get video streams from other sites and display them on screens. It has special software to manage the video streams, runs Windows 2000 operating system, and has a multi-headed video card. The video capture computer is used to get the video streams from the cameras in the room. It has four video capture cards and runs Linux operating system. The audio capture computer gets audio streams from the microphones in the room and encodes and broadcasts them to other nodes. It also gets audio streams from remote nodes and decodes them. It runs Linux operating system. The control computer is used to run control software for the audio gear (echo canceller). It runs Windows 98 operating system.

3.6 Software

Access Grid partners have developed several pieces of software. One is a multicast beacon used to monitor the network status of nodes. Another is a distributed PowerPoint tool used to share PowerPoint slides in a session. Persistence and scope are provided by using the Virtual Venue software developed at Argonne. It has components that run on the display, video, and audio machines, as well as a central server. Video Conferencing Tool (VIC) is another software used to manage displays, and Robust Audio Tool (RAT) software is used to manage audio.

3.7 Network

The access grid uses network multicast among AG nodes. A full AG session can deliver several dozen video streams to a node. The bandwidth required for each stream can vary from 128 Kb/s to 512Kb/s depending on the settings. Inadequate bandwidth results in unintelligible audio and jerky-motion video.

3.8 Protocols

The Access Grid uses RAT, an open source software, for handling audio. It is an audio conferencing and streaming application that allows users to participate in audio conferences over the Internet. RAT is based on IETF standards and uses RTP above UDP/IP as its transport protocol. RAT features a range of different rate and quality codecs, G.711(64kb/s), Wide-Band ADPCM(64kb/s), G.726 ADPCM (16-40kb/s), DVI ADPCM (32kb/s), Variate Rate DVI ADPCM (~32kb/s), Full Rate GSM (13kb/s), and LPC (5.6kb/s). It also features encryption for private conversations.

The Access Grid uses VIC to handle video. VIC is a real-time, multimedia application for video conferencing over the Internet. It was developed by Network Research Group at the Lawrence Berkeley National Laboratory in collaboration with the University of California, Berkeley. VIC is based on Real Time Transport Protocol (RTP) developed by IETF. To use VIC's conferencing capabilities, systems should support IP multicast. VIC uses H.261 protocol to encode and decode video streams. H.261 is the protocol that defines the video portion of H.323.

3.9 Recording/Playback

Argonne has built a recording and playback engine, Voyager Multimedia Multistream, that can record and playback live sessions. It saves multiple video and audio streams to disks without loss. It also synchronizes in time the multiple audio and video streams when playing back.

3.10 Required Equipment

An Access Grid node consists of several pieces of hardware equipment:

- Four PCs
  - Display computer
  - Video capture computer
  - Audio capture computer
  - Control computer
- Four cameras
- Several microphones
- Echo canceller device
- Three projectors or displays

3.11 Cost

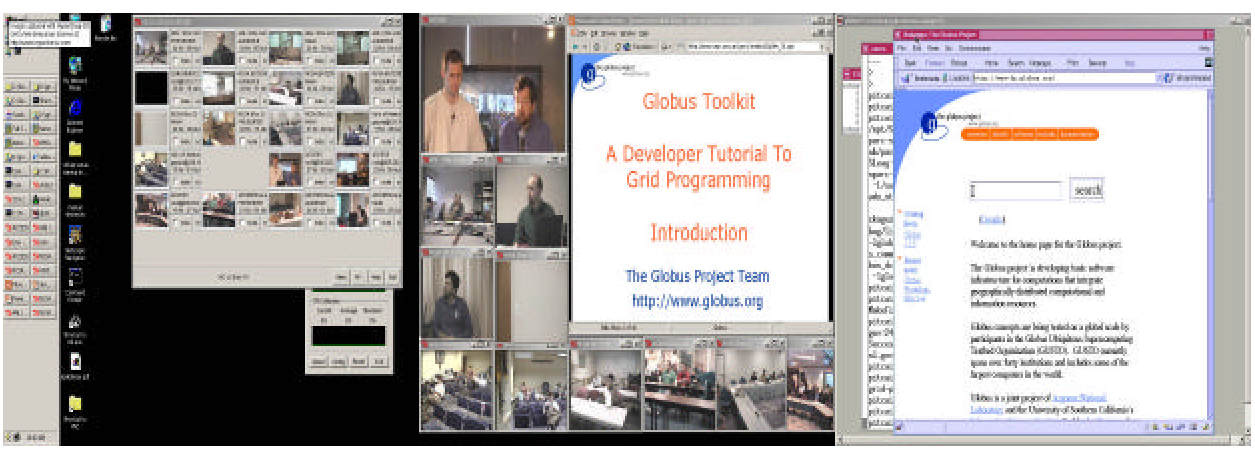
Computing equipment	$12,455
Network equipment	$750
Other computing equipment (monitors, KVM switch)	$1,800
Audio configuration	$10,564
Video cameras (four Sony EVI-D30)	$5,196
Projectors (three Epson 710c)	$15,900
Total (January 2001)	$46,665

These prices and equipment may vary depending on the configuration of the AG node. Access Grid software is free and will be available on a CD.

3.12 Conclusion

Today, group-to-group collaboration is needed in many areas since it is not easy to gather everyone in the same place. Access Grid offers a solution to collaborating with remote locations by providing real, lifesize images and hands-free audio. This has proved successful, and the number of institutions that are installing the Access Grid is increasing rapidly.

Here are some photographs from an Access Grid session:



4 References

1) Geoffrey C. Fox, "Architecture and Implementation of a Collaborative Computing and Education Portal," ERDC technical report, May 2001.
2) HearMe http://www.hearme.com/
3) Access Grid http://www-fp.mcs.anl.gov/fl/accessgrid/default.htm
4) FSU Review of Collaboration Tools http://aspen.csit.fsu.edu/collabtools/CollabReviewmay09-01.doc
5) AES Encryption Standard http://csrc.nist.gov/encryption/aes/